

Name: Faiz Yah



**1. Own definition of data science based on own understanding.**

- When we mention about scientists, the first impression that comes to many people's minds would be of a chemist, mixing and experimenting different chemicals and substances together to form new formulas. Taking a similar analogy, data science is essentially taking raw data, massaging and processing it, experimenting with various methods and approaches to obtain impactful insights, translating raw numbers into meaningful information.

**2. Identify famous data scientist.**

- As data scientist, an incredible and widely used library is **scikit-learn**, which was initially developed by David Cournapeau that is originally from France. As someone who recently have a newfound passion with the French language, I realised more and more a sizeable amount of contributors in the Data and AI field are of French origin, including **Mistral AI** founder, Arthur Mensch, **Keras** deep-learning framework by François Chollet, to name a few. Andrew Ng is without a doubt another famous data scientist considering the immense impact he has contributed with his research.

**3. List domains of DS applications and examples of data products.**

- Data science applications can be found in any and every domain imaginable. The only prerequisite for data science is data. As long as there is data, we can apply data science principles and methodologies to the data to arrive at our goal, either predicting analytics, obtaining insights for informed-decision, or even as simple as cleaning our data. Few popular domains of data science application include healthcare domain for diseases

early detection, finance domain for credit scoring and fraud detection, social media domain to build recommender system through graph data science, marketing domain using customer sentiment analysis for a certain product and many more.

**4. Rationale why data science matters today.**

- The relevancy and demand of data science today is stronger than ever and is only expected to continue with this incrementing trajectory. One solid reason for this is the data that each human generates every second is also increasing. As more and more data is available, the impactful use-cases and possibilities from leveraging these data increases as well. The positive impact of data science is also prevalent and can be virtually felt by everyone, from using Google Maps in helping us optimizing the best route with the least amount of travel time, to having personalised shopping experience on Shopee, these are all contributions of Data Science. The magnitude of this importance is further heightened by the increasing discovery and breakthrough in this space, including better transformers, improved deep learning techniques, smarter GenAI and AI agents.

**5. Provide methodology, process, lifecycle, or workflow for data science.**

- A typical lifecycle for data science starts from understanding the business needs and problem statement. Data is then collected based on the context. Data collected are usually in a raw format and is uncleaned, so data cleaning and processing is the next step. Once data is clean, we proceed with exploratory data analysis to uncover insights and understand our data better. If necessary, more relevant features are created through feature engineering, which requires strong domain knowledge. A model is then trained and tested in an iterative process, each time tuning the hyperparameters to identify the most optimal one. Finally, the results and outcomes are presented in a carefully crafted report with powerful data visualisations, storytelling here is vital to ensure the right narrative and data is translated across.